

UNITED STATES PATENT APPLICATION

OF

**WILLIAM J. BUISMAN
GARY COHEN
STEVE R. FOX
CHARLES KRUSE
CHRISTINE SULPIZIO**

FOR

SYSTEM AND METHOD FOR CONVERTING INFORMATION ON PAPER FORMS TO ELECTRONIC DATA

DECEMBER

**SYSTEM AND METHOD FOR CONVERTING INFORMATION ON PAPER
FORMS TO ELECTRONIC DATA**

CROSS-REFERENCE TO RELATED APPLICATIONS

5

The subject matter of this application is related to the subject matter of provisional U.S. Patent Application Serial No. 60/182,674 filed February 15, 2000 entitled "SYSTEM AND METHOD FOR PROCESSING AND TRACKING APPLICATIONS FOR FINANCIAL PRODUCTS AND SERVICES," which application
10 is assigned or under obligation of assignment to the same assignee as this application and which is incorporated by reference herein, and priority being claimed therefrom.

FIELD OF THE INVENTION

The invention relates generally to a system and method for processing data, and
15 more specifically for converting information on paper forms to electronic data which can be utilized by subsequent processes.

BACKGROUND OF THE INVENTION

Many commercial and government entities receive input data in paper format.
20 Credit card applications, license applications, and tax returns are examples. Although input techniques such as automated telephone systems and Web-based utilities are increasingly provided as an input alternative, the volume of input made in paper format is still substantial. For instance, a bank or other financial institution may receive more than

10,000 credit card applications in paper format in a single day. Manual entry of this volume of data into data processing systems may be cost prohibitive.

5 A common technique for processing such a large volume of input forms involves automated or semi-automated conversion of the completed paper forms to electronic format. In a typical approach, paper forms are first scanned by a digital image scanner to yield an electronic bitmap image. The image is then converted to text via Optical Character Recognition (OCR) software (for reading machine printed characters) and Intelligent Character Recognition (ICR) software (for reading hand written characters).
10 Not all data is correctly interpreted, however, due to functional limitations in scanning apparatus and recognition software. It is therefore common practice to employ data entry operators for the purpose of correcting errors or omissions resulting from the automated conversion process. For an example of such a system and method, see U.S. Pat No. 5,054,096 issued to Beizer on Oct. 1, 1991.

15 Known approaches have several drawbacks and limitations, however. One issue not addressed by existing technology is how to manage the cost and cycle time associated with data entry staff. Another limitation is a failure to recognize that it may be advantages to process some applications differently than others, according to a variety of factors.

20 In sum, existing systems and techniques for converting information on paper forms to electronic data have not adequately managed the conversion process. The resulting lack of efficiency, and other drawbacks, limit the utility of such systems for entities receiving a high volume of input data in paper format.

SUMMARY OF THE INVENTION

The invention overcoming these and other drawbacks in the art relates a system and method for converting credit card applications and other forms to electronic format.

5 The system and method may operate on full forms or only portions (or snippets) of the forms, which are later reassembled. In one embodiment, conversion processing may be accomplished through the selective use of internal and external data entry operators. Work may also be prioritized according to its importance to the processing entity.

10 It is an object of the invention to reduce the cost and cycle time associated with the conversion process from paper format to electronic data format.

It is another object of the invention to enable an entity using the conversion process to tailor the work flow to the needs of their organization.

15 The following drawings and descriptions further describe the invention, including different embodiments of the major system components and processes. The construction of such a system, implementation of such a process, and advantages will be clear to a person skilled in the art of document conversion.

BRIEF DESCRIPTION OF THE DRAWINGS

20 Figure 1 is schematic diagram of a system configured for the conversion of paper forms into electronic data, according to one embodiment of the invention.

Figure 2 is a high-level flow diagram for forms-based processing, according to one embodiment of the invention.

Figure 3 is a more detailed flow diagram of form preparation, according to one embodiment of the invention.

Figure 4 is a more detailed flow diagram of data capture, according to one embodiment of the invention.

Figure 5 is a schematic diagram illustrating how a form is parsed into snippets,
5 according to one embodiment of the invention.

DETAILED DESCRIPTION OF THE INVENTION

Figure 1 is schematic diagram of the system, according to one embodiment of the invention. The diagram illustrates an overall architecture for conversion processing, including: paper forms 100, 102, and 114; a facsimile machine 112; image scanners 104,
10 106, and 116; servers 128, 132, 134, 136, and 138; database 130; clients 108, 110, 118, 120, 122, 124, and 126; and communication links 140, 141, 142, and 143.

Paper forms 100 and 102 may be received via an intake service, such as a mail delivery service. Paper form 114 is the output from facsimile machine 112. Paper forms 100, 102, and 114 may be imaged by digital image scanners 104, 106, and 116 or other
15 image generator. Optical Character Recognition (OCR) and Intelligent Character Recognition (ICR) software, which may be resident on imaging devices 104, 106, and 116 or on input clients 108, 110, and 118, may convert the images into alpha-numeric text. The system may include internal data entry clients 120 and 122, and may also include external data entry clients 124 and 126.

20 Clients 108, 110, 118, 120, 122, 124, and 126 may be or include, for instance, a personal computer running the Microsoft Windows™ 95, 98, Millenium™, NT™, or 2000, Windows™CE™, PalmOS™, Unix, Linux, Solaris™, OS/2™, BeOS™, MacOS™ or other operating system or platform. Clients 108, 110, 118, 120, 122, 124, and 126 may include a microprocessor such as an Intel x86-based device, a Motorola 68K or

PowerPC™ device, a MIPS, Hewlett-Packard Precision™, or Digital Equipment Corp. Alpha™ RISC processor, a microcontroller or other general or special purpose device operating under programmed control. Clients 108, 110, 118, 120, 122, 124, and 126 may
5 furthermore include electronic memory such as RAM (random access memory) or EPROM (electronically programmable read only memory), storage such as a hard drive, CDROM or rewritable CDROM or other magnetic, optical or other media, and other associated components connected over an electronic bus, as will be appreciated by persons skilled in the art. Clients 108, 110, 118, 120, 122, 124, and 126 may also be or
10 include a network-enabled appliance such as a WebTV™ unit, radio-enabled Palm™ Pilot or similar unit, a set-top box, a networkable game-playing console such as Sony Playstation™ or Sega Dreamcast™, a browser-equipped cellular telephone, or other TCP/IP client or other device.

Server 128 may control access to database 130, and may also control work flow
15 between servers 132, 134, 136, and 138. Servers 128, 132, 134, 136, and 138 may be or include, for instance, a workstation running the Microsoft Windows™ NT™, Windows™ 2000, Unix, Linux, Xenix, IBM AIX™, Hewlett-Packard UX™, Novell Netware™, Sun Microsystems Solaris™, OS/2™, BeOS™, Mach, Apache, OpenStep™ or other operating system or platform.

20 Clients 108, 110, 118, 120, 122, 124, and 126, and servers 128, 132, 134, 136, and 138 may utilize network enabled code to exchange data or instructions over communications links 140, 141, 142, and 143. The network enabled code may be, include or interface to, for example, Hyper text Markup Language (HTML), Dynamic HTML, Extensible Markup Language (XML), Extensible Stylesheet Language (XSL),

Document Style Semantics and Specification Language (DSSSL), Cascading Style
Sheets (CSS), Synchronized Multimedia Integration Language (SMIL), JavaTM, JiniTM,
C, C++, Perl, UNIX Shell, Visual Basic or Visual Basic Script, Virtual Reality Markup
5 Language (VRML) or other compilers, assemblers, interpreters or other computer
languages or platforms.

Communications links 140, 141, 142, and 143 may be, include or interface to any
one or more of, for instance, the Internet, an intranet, a PAN (Personal Area Network), a
LAN (Local Area Network), a WAN (Wide Area Network) or a MAN (Metropolitan
10 Area Network), a storage area network (SAN), a frame relay connection, an Advanced
Intelligent Network (AIN) connection, a synchronous optical network (SONET)
connection, a digital T1, T3, E1 or E3 line, Digital Data Service (DDS) connection, DSL
(Digital Subscriber Line) connection, an Ethernet connection, an ISDN (Integrated
Services Digital Network) line, a dial-up port such as a V.90, V.34 or V.34bis analog
15 modem connection, a cable modem, an ATM (Asynchronous Transfer Mode) connection,
or an FDDI (Fiber Distributed Data Interface) or CDDI (Copper Distributed Data
Interface) connection. Communications links 140, 141, 142, and 143 may furthermore
be, include or interface to any one or more of a WAP (Wireless Application Protocol)
link, a GPRS (General Packet Radio Service) link, a GSM (Global System for Mobile
20 Communication) link, a CDMA (Code Division Multiple Access) or TDMA (Time
Division Multiple Access) link such as a cellular phone channel, a GPS (Global
Positioning System) link, CDPD (cellular digital packet data), a RIM (Research in
Motion, Limited) duplex paging type device, a Bluetooth radio link, or an IEEE 802.11-
based radio frequency link. Communications links 140, 141, 142, and 143 may yet

further be, include or interface to any one or more of an RS-232 serial connection, an IEEE-1394 (Firewire) connection, a Fibre Channel connection, an IrDA (infrared) port, a SCSI (Small Computer Systems Interface) connection, a USB (Universal Serial Bus) connection or other wired or wireless, digital or analog interface or connection.

The database 130 may be, include or interface to, for example, the OracleTM relational database sold commercially by Oracle Corp. Other databases, such as InformixTM, DB2 (Database 2) or other data storage or query formats, platforms or resources such as OLAP (On Line Analytical Processing), SQL (Standard Query Language), a storage area network (SAN), Microsoft AccessTM or others may also be used, incorporated or accessed in the invention.

Figure 2 is a high-level flow diagram for processing of financial applications or other forms, according to one embodiment of the invention. The illustrated process begins with form preparation 200, which is further described by Figure 3. Form distribution 210 may involve, for example, delivery of forms to existing or potential customers. Form distribution 210 may be accomplished by direct mail, by placement on countertop or rack, or by other suitable procedure. Form completion 220 may be satisfied with handwritten data entry onto the form, by using a pen or pencil, for example. Form completion 220 may also involve machine-aided data entry such as use of a typewriter or electromechanical printer. The data capture process 230 converts data added to the form into electronic format, and is further illustrated in Figure 4. The data capture process 230 may be internal to an organization, and may be supplemented with external data entry 240. The type and frequency of errors related to data capture process 230 may be continuously monitored or periodically audited by a quality control function

250. One use of the data capture process 230 may be to enable transaction processing
260, such as automated review of credit card applications. The output of data capture
process 230 may also be exploited by a data storage process 270 to create a useful
5 database.

Figure 3 is a more detailed flow diagram of form preparation 200, according to
one embodiment of the invention. The diagram illustrates that a new form is created 310
only after determination of a need 300. The form creation process 310 may receive an
assigned identification number 320 as an input. Identification number 320 may appear
10 on the resulting form as printed text, or it may be encoded in machine-readable format
such as a barcode. The form creation process 310 may be further constrained by certain
content and format requirements 330. A content requirement might be, for example, that
all forms include the name and address of the form provider and request the name and
address of the person completing the form. Format restrictions may include requirements
15 such as font type, minimum character size, and line spacing. Before the form preparation
step 200 is considered complete, an entity may require that a sample form be tested for
readability 340, for example with existing scanner hardware and character recognition
software.

Figure 4 is a more detailed flow diagram of data capture process 230, according
20 to one embodiment of the invention. The process starts 400 with receipt of a completed
form. The first step is to scan or otherwise read the form 402, converting the form and
all data contained thereon into electronic format. This may be accomplished via digital
image scanners 104, bar code readers, Optical Character Recognition (OCR) software,

Intelligent Character Recognition (ICR) software, or through similar reading techniques known to data conversion practitioners.

Processing may be contingent on the outcome 404 of the read 402. A form code
5 identifies the particular type of form. If the form code was successfully interpreted, then
a priority may be assigned to the form 410. Assignment of priority may be based, for
example, on a preferred client list, according to the profit margin that a seller of goods or
services expects to receive from a buyer, or according to other criteria. The priority
assigned in step 410 may be represented on a scale of 1 to 10, designated by high,
10 medium, or low, or otherwise rank ordered. If the form code read was not successful in
step 402, then the form may be routed to data repair 406 for manual classification of the
incoming form type.

Work flow may be further contingent on whether data repair 406 was successful
408. If data repair 406 was a success, for example where the form code was human-
15 readable, then the form is promoted to step 410 for assignment of priority. If, on the
other hand, data repair 406 was not successful in determining the form code, then the
form may be designated as an unknown form type 414.

After forms have been assigned a priority 410, they may be reviewed for change
of address 412. Change of address may be detected by a box that has been checked, for
20 example, or by the presence of text outside of defined data input areas. Where a change
of address has not been detected, the form may be routed by decision process 412 to
parsing process 416. An unknown form 414, or a form with a change of address 412,
may be processed as a full image 418.

Parsing 416 decomposes form data into snippets. Figure 5 illustrates a form 501 that has been parsed into a Name Snippet 502, an Address Snippet 503, and a Social Security Number Snippet 504. Subsequent processing of parsed data is advantageous because a majority of snippet types are common to different form types. The result is that new form types are easily introduced into the semi-automated data capture process 230. Another advantage of snippet processing is information security. For instance, a data entry operator verifying data originally contained on form 501 is not able to associate the social security number with a name: the data entry operator is merely processing a series of unassociated name and social security number snippets.

The daily volume of forms received 400 for data capture 230 may vary substantially. For this and other reasons, it may be advantageous to utilize external data entry vendors 240 capable of processing either parsed 416 or full image forms 418. Figure 4 therefore indicates that electronic form data may be transmitted 420 to external data entry vendors for the purpose of verifying the automated read 402. External data entry 240 may involve, for example, an on-screen comparison between a bit mapped image of a snippet or full image and the textual equivalent produced by character recognition software. After external data entry 240, the system may receive 422 snippets or full images from the external data entry vendor. In the case of snippet processing by external data entry vendors 240, the data is repackaged 424 so that snippets are re-associated according to original format received in step 400.

When form data is received from external data entry vendors 422, and repackaged 424 if necessary, a decision 426 may be made as to the need for review by internal data entry operators 428. Such a review may be appropriate, for example, where data is

missing or could not be discerned by operators employed by the external data entry operator. If no review is needed, the verified electronic forms may be sent to transaction processing 260 or data storage 270.

5 Finally, Figure 4 also illustrates that the order in which form data is operated on by the transmit 420, repackage 424, and internal data entry 428 processes may be dictated by the priority assigned in step 410. Processing in step 428 may also be prioritized or sorted according to the type of errors that could not be resolved by external data entry vendors in step 240. For instance, all errors resulting from missing data might be
10 processed in step 428 only by internal data entry operators who are trained and equipped to reach applicants by telephone.

 The specification and examples provided above should be considered exemplary only. It is contemplated that the appended claims will cover any other such embodiments or modifications as fall within the true scope of the invention.